An Engineering Perspective on Good Practice

Alex Ball

24 April 2013

Abstract

Engineering is a discipline without a funder template for data management plans (DMPs). How does one go about writing a DMP in such circumstances? The REDm-MED Project put together data management procedures for the Department of Mechanical Engineering at the University of Bath. The major challenges included confidentiality of data, heterogeneity of data and incomplete data management infrastructure.

I've been asked to speak to you about Data Management Planning in the Department of Mechanical Engineering at Bath. Engineering…not much data sharing. In 2005–7, a project called PerX (`http://www.engineering.ac.uk/`) developed an Engineering Repository Cross Search Demonstrator (¶ Figure 1), but while they could find technical reports, theses, papers and the odd learning object, they found very little in the way of research data.

Nevertheless the mood is shifting towards more data sharing. ¶ In disciplines like Engineering, it will no longer be taken as read that data cannot be shared: researchers will have to provide clear justification. The EPSRC says it expects as much (here's the full list of expectations, paraphrased):

1. *Research organisations (ROs)* to raise awareness of data sharing responsibilities and issues.
2. Publications should link to underlying data.
3. *ROs* must keep track of their research datasets and requests for them.
4. Born-analogue data must also be shareable on request.
5. *ROs* must provide open, online catalogues of their data; digital data must be given a robust ID.
6. Access restrictions should be clear and justified.
7. *ROs* must provide access to data for 10 years from last access.
8. *ROs* must curate their research data.
9. *ROs* must pay for this from their existing public funding streams.

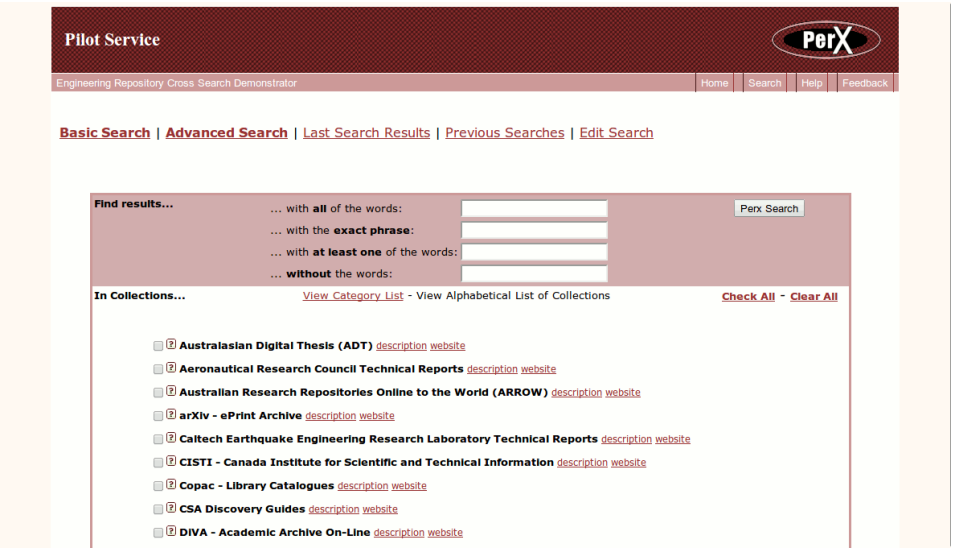


Figure 1: PerX

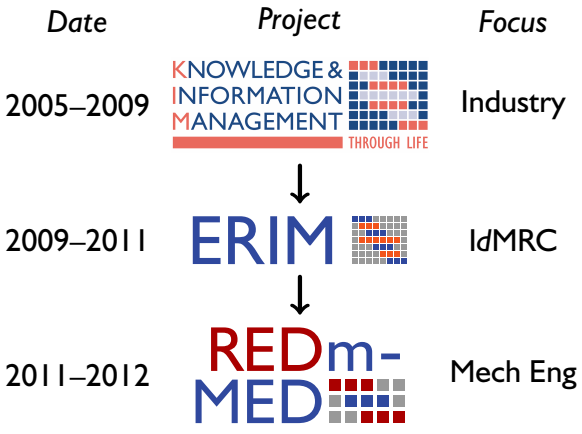


Figure 2: Data management projects at the Department of Mechanical Engineering at the University of Bath

The engineers at Bath have had a long interest in data management (¶ Figure 2).

For me it started with the KIM Project…collaboration between 11 universities …interested in information organisation, selection policies for data, and finding ways of learning design lessons from in-service data and knowledge.

In ERIM, we tried to implement a data management planning regime at the Innovative Design and Manufacturing Research Centre. We looked at the KIM data and other data being produced by the Centre, and at how well they had been managed so far. ¶

We found, unsurprisingly, that the Centre, and by extension the Department and the University, had a…

1. Poor framework for
    - pre-project considerations of data management; *no tools or guidance for DMPs*
    - data management during the research; *only shared drives and a few conventions: project data folders, plain text indices, 7-year retention policy from project end*
    - during-project data management for post-project re-use. *no collection of metadata*
2. Poor knowledge of context in which data were generated:
    - engineering research data is very diverse; *photos, videos, process flow diagrams, materials data sheets, activity models, product models, topic maps, bills of materials, questionnaire responses, interview transcripts, 3D laser scans, flight path data, thermal data profiles, surface roughness data…*
    - large number of diverse research data records; *lots of little ones, hardly any big ones*
    - relations between data records complex. *have videos of researchers explaining the data they hold, not easy to understand at all, turned out vital records missing …*
3. Knowing the context is vital for understanding data. *Slide: Mansur Darlington*

So this got us thinking about what we needed our data management framework to do, and being so inclined we decided to derive it from first principles (¶ Figure 3).

The *Principles of Engineering Information Management* came out of the KIM Project. 11 principles…inspired by good practice and lessons learned from industry, as well as digital curation theory ¶:

1. Parsimony – *Create, record and retain information only if necessary*
2. Granularity – *Record information in a storable information object at a granularity appropriate for use and re-use*
3. Identity – *Give an information object a unique and persistent identifier*

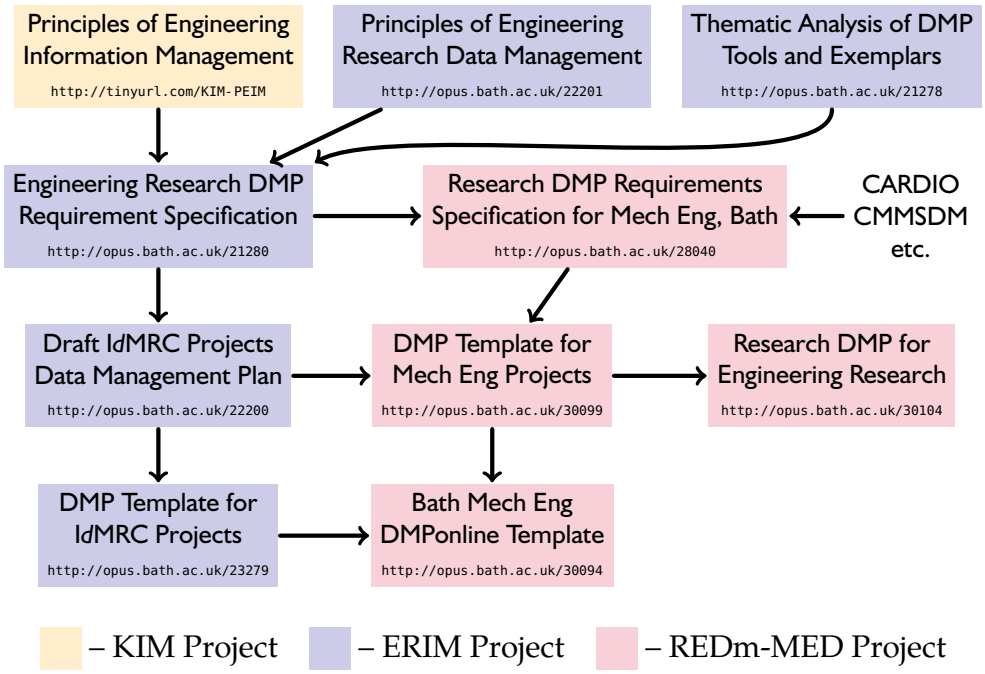


Figure 3: Engineering research data management planning guidance (simplified!)

4. Uniqueness – *Create an information entity once only and explicitly reference it everywhere else*
5. Usability – *Design an information entity explicitly to achieve its intended goals*
6. Reusability – *Design an information entity explicitly to maximise its potential for reuse wherever appropriate*
7. Evaluation – *Assess and assign the value of an information object throughout its life from creation to disposal*
8. Portability – *Create an information entity and its annotations systematically using representations supporting perpetual re-use*
9. Robustness – *Use robust methods to capture, create and manipulate information entities*
10. Discovery – *Actively employ the information repository as a resource for learning and discovery*
11. Design – *Design all aspects of information management to satisfy the organisation's current and future needs*

¶§ In the ERIM project we had to bring the focus round to research data management, which we did with two documents: we looked at the DMP guidance already out there and picked out what issues they were getting at, and produced a complementary set of RDM principles ¶.

1. See the Principles of Engineering Information Management.
2. See the DCC Charter and Statement of Principles – *e.g. Promote the practice of creating documentation and metadata as a means of providing context for datasets, in order to facilitate the future discovery, access, use and reuse of data.*

3. Data processing should be **reproducible**.
4. Use **generic/standard tools** where possible.
5. DMPs should support repurposing and help to support data reuse.
6. Treat records of a research activity **together as a set**.
7. Make **context/associations** between records explicit.
8. Methods of recording context should also be documented.
9. Confidentiality agreements should be as permissive as ethically possible.
10. RDM should be **costed** into each research proposal.
11. Supporting reuse/repurposing should **not get in the way** of use.
12. Any RDM tools should be **simple, engaging & easy to access**.

¶ From these we derived a requirements specification for engineering research DMPs.

¶ At this point we were only worried about the reusability of data and set aside some important issues like long-term preservation.

Infrastructure

1. Links from DMP to key docs
2. Links from key docs to DMP
3. Purpose of DMP
4. Roles and responsibilities
5. Review and adherence
6. Version control
7. Budget
8. Storage, backup and security
9. Receiving repository

DMP contents

1. Summary of activity
2. Reuse of existing data
3. Fitting in with existing data
4. Preparing for expected reuse
5. Record manifest
6. Data generation and manipulation
7. Data organisation
8. Quality assurance
9. Data structures and formats
10. Data semantics

For each of these we stated a *requirement* and gave *guidance* and *rationale*. We then fleshed out what some of the answers should be in the case of the I*d*MRC, and eventually produced a template that researchers could fill in.

One of the things we were particularly concerned about was the record manifest, in the light of our principle about recording context. We had an idea for a tool that researchers could use to record how their data assets were related, but didn't have resources to write it. ¶ So we received funding for a short project called REDm-MED to develop this tool, extend and implement our ERIM work department-wide, and package all this up so other departments like yourselves could use them too.

¶§ We updated our requirements in the light of a CARDIO assessment of the Mech Eng Department, and emerging work such as the Capability Maturity Model for Scientific Data Management (`http://dx.doi.org/10.1002/meet.2011.14504801036`). ¶ While the ERIM requirements specification was simply *requirement*, *guidance*, *rationale*, with REDm-MED we went a little further and gave:

- Requirement
- Rationale
- Validation (DCC Checklist, applicability to research and data lifecycles)
- Role supported by requirement
- Responsibility – institution, department or project?
- Information/resource needed – *we provided a separate document that listed all of these and explained what they were all for:* `http://opus.bath.ac.uk/29583`

We produced new versions of the data management procedures and plan template, and wrote a DMPonline Template that our researchers could fill in. I'll conclude by giving you a flavour of what we proposed ¶.

Should be able to find quickly:

- Data management plan/record
- Project proposal (pre-award)
- Detailed project plan (post-award)
- Project record manifest
- Confidentiality agreements – *as long as they are not confidential!*
- IPR statements, licences – *strongly affect how data may be reused*

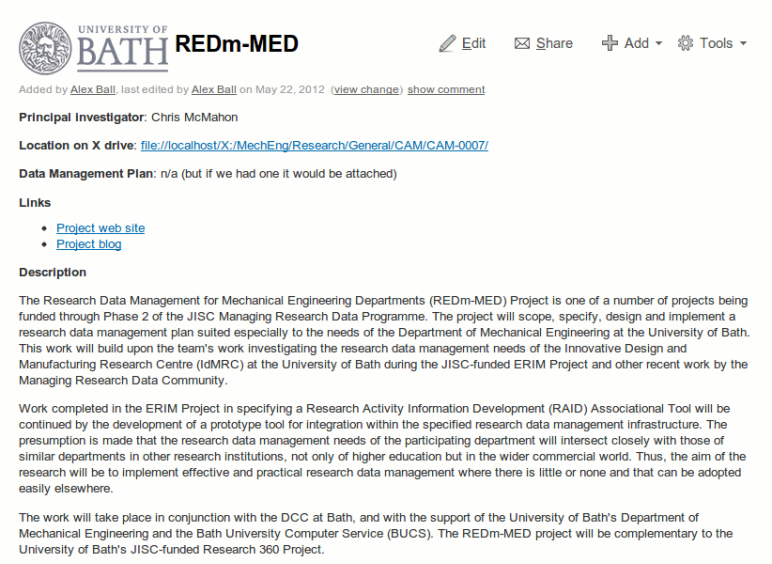
To make this possible, we set up a Departmental Research Data Management wiki (`http://wiki.bath.ac.uk/display/MechEngRDM/`), where every project is supposed to put a page listing key locations (¶ Figure 4). The idea is that anything public goes there. If there are private documents, they are kept in an access-controlled area (perhaps an intranet); this public page links to that private space but obviously doesn't tell you how to break in.

The other thing of interest I should mention is the Project Record Manifest. This is a list of all the data records produced by a research activity. We provided two ways of assembling one; the first was a project data record list (¶ Figure 5).

The main component of this (let me zoom in §) is a table listing all the records associated with a project, showing the record title, file name and location, owner and contact details, record type, and confidentiality status.

The second (¶ Figure 6) was to use the RAIDmap application (`http://sourceforge.net/p/raidmap/`) to create a map of associations between records.

So, if you are developing your own DMP regime, what sort of questions do you need to ask yourself?



UNIVERSITY OF BATH **REDm-MED**

Edit Share Add Tools

Added by Alex Ball, last edited by Alex Ball on May 22, 2012 (view change) show comment

Principal Investigator: Chris McMahon

Location on X drive: file://localhost/X:/MechEng/Research/General/CAM/CAM-0007/

Data Management Plan: n/a (but if we had one it would be attached)

Links

- Project web site
- Project blog

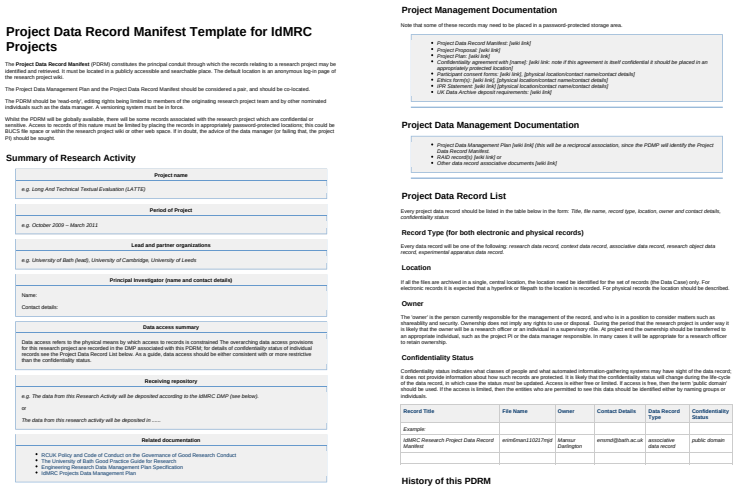
Description

The Research Data Management for Mechanical Engineering Departments (REDm-MED) Project is one of a number of projects being funded through Phase 2 of the JISC Managing Research Data Programme. The project will scope, specify, design and implement a research data management plan suited especially to the needs of the Department of Mechanical Engineering at the University of Bath. This work will build upon the team's work investigating the research data management needs of the Innovative Design and Manufacturing Research Centre (IdMRC) at the University of Bath during the JISC-funded ERIM Project and other recent work by the Managing Research Data Community.

Work completed in the ERIM Project in specifying a Research Activity Information Development (RAID) Associational Tool will be continued by the development of a prototype tool for integration within the specified research data management infrastructure. The presumption is made that the research data management needs of the participating department will intersect closely with those of similar departments in other research institutions, not only of higher education but in the wider commercial world. Thus, the aim of the research will be to implement effective and practical research data management where there is little or none and that can be adopted easily elsewhere.

The work will take place in conjunction with the DCC at Bath, and with the support of the University of Bath's Department of Mechanical Engineering and the Bath University Computer Service (BUCS). The REDm-MED project will be complementary to the University of Bath's JISC-funded Research 360 Project.

Figure 4: Departmental Research Data Management Wiki



Project Data Record Manifest Template for IdMRC Projects

The **Project Data Record Manifest** (PDRM) constitutes the principal conduit through which the records relating to a research project may be identified and retrieved. It must be located in a publicly accessible and searchable place. The default location is an anonymous log-in page of the research project wiki.

The Project Data Management Plan and the Project Data Record Manifest should be considered a pair, and should be co-located.

The PDRM should be 'read-only', editing rights being limited to members of the originating research project team and by other nominated individuals such as the data manager. A versioning system must be in force.

Whilst the PDRM will be globally available, there will be some records associated with the research project which are confidential or sensitive. Access to records of this nature must be limited by placing the records in appropriately password-protected locations; this could be BUCS file space or within the research project wiki or other web space. If in doubt, the advice of the data manager (or failing that, the project PI) should be sought.

Summary of Research Activity

Project name
e.g. Long And Technical Textual Evaluation (LATTE)
Period of Project
e.g. October 2009 – March 2011
Lead and partner organizations
e.g. University of Bath (lead), University of Cambridge, University of Leeds
Principal Investigator (name and contact details)
Name: Contact details:
Data access summary
Data access refers to the physical means by which access to records is constrained The overarching data access provisions for this research project are recorded in the DMP associated with this PDRM; for details of confidentiality status of individual records see the Project Data Record List below. As a guide, data access should be either consistent with or more restrictive than the confidentiality status.
Receiving repository
e.g. The data from this Research Activity will be deposited according to the IdMRC DMP (see below). *or* *The data from this research activity will be deposited in .....*
Related documentation
• RCUK Policy and Code of Conduct on the Governance of Good Research Conduct • The University of Bath Good Practice Guide for Research • Engineering Research Data Management Plan Specification • IdMRC Projects Data Management Plan

Project Management Documentation

Note that some of these records may need to be placed in a password-protected storage area.

- *Project Data Record Manifest: [wiki link]*
- *Project Proposal: [wiki link]*
- *Project Plan: [wiki link]*
- *Confidentiality agreement with [name]: [wiki link: note if this agreement is itself confidential it should be placed in an appropriately protected location]*
- *Participant consent forms: [wiki link], [physical location/contact name/contact details]*
- *Ethics form(s): [wiki link], [physical location/contact name/contact details]*
- *IRR Statement: [wiki link] [physical location/contact name/contact details]*
- *UK Data Archive deposit requirements: [wiki link]*

Project Data Management Documentation

- *Project Data Management Plan [wiki link] (this will be a reciprocal association, since the PDMP will identify the Project Data Record Manifest.*
- *RAID record(s) [wiki link] or*
- *Other data record associative documents [wiki link]*

Project Data Record List

Every project data record should be listed in the table below in the form: *Title, file name, record type, location, owner and contact details, confidentiality status*

Record Type (for both electronic and physical records)

Every data record will be one of the following: *research data record, context data record, associative data record, research object data record, experimental apparatus data record.*

Location

If all the files are archived in a single, central location, the location need be identified for the set of records (the Data Case) only. For electronic records it is expected that a hyperlink or filepath to the location is recorded. For physical records the location should be described.

Owner

The 'owner' is the person currently responsible for the management of the record, and who is in a position to consider matters such as retention and disposal. Ownership does not imply any rights to use or dispose. During the period that the research project is underway it is likely that the owner is a member of the research team or an individual in a supervisory role. As project end the ownership should be transferred to an appropriate individual, such as the project PI or the data manager responsible. In many cases it will be appropriate for a research officer to retain ownership.

Confidentiality Status

Confidentiality status indicates what classes of people and what automated information-gathering systems may have sight of the data record; it does not provide information about how such records are protected. It is likely that the confidentiality status will change during the life-cycle of the data record, in which case the status must be updated. Access is either free or limited. If access is free, then the term 'public domain' should be used. If the access is limited, then the entities who are permitted to see this data should be identified either by naming groups or individuals.

Record Title	File Name	Owner	Contact Details	Data Record Type	Confidentiality Status
Example					
IdMRC Research Project Data Record Manifest	*erim6mar110217mjd*	*Mansur Darlington*	*emsmd@bath.ac.uk*	*associative data record*	*public domain*

History of this PDRM

Figure 5: Project Data Record Manifest Template for IdMRC Projects

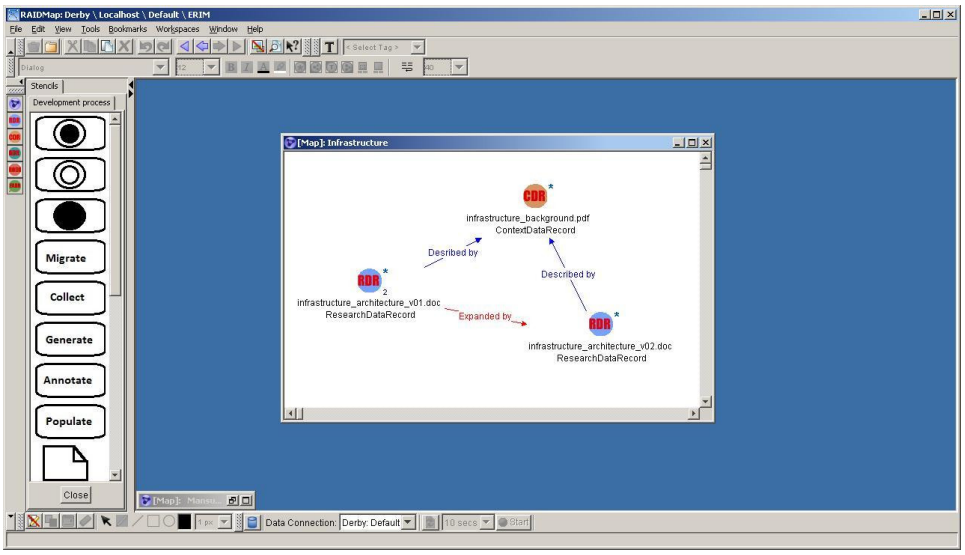


Figure 6: RAIDmap

- What do you want to achieve? – *compliance with EPSRC expectations would be a good start!* `http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx`
- What can be done by the institution/department, and what is left to researchers? – *arguably only the stuff by researchers needs to go into the written plan.*
- What can be done with what you already have, and what do you need to improve? – *we could get a fair way with our wiki and shared drive, but need an institutional repository; Research360 is working on it.*
- Can you introduce research data management so it decreases the burden on researchers? – *if you force people to do it, they'll do it grudgingly and badly; focus on benefits that will make them want to do it properly. This is the hardest bit of the process!*

Alex Ball. DCC/UKOLN, University of Bath. `http://www.ukoln.ac.uk/ukoln/staff/a.ball/`

ERIM `http://www.ukoln.ac.uk/projects/erim/`

REDm-MED `http://www.ukoln.ac.uk/projects/redm-med/`

Research360 `http://blogs.bath.ac.uk/research360/`